DIGITAL SIGNAL PROCESSING LABORATORY

EXPERIMENT-5 – TEMPORAL SPEECH RECOGNITION

Group:- 32

Authors:-

Jaya Kishnani (20EC30020)

Gunjan Shekhar(20EC10032)

TA:- Abhishek Singh

Aim:

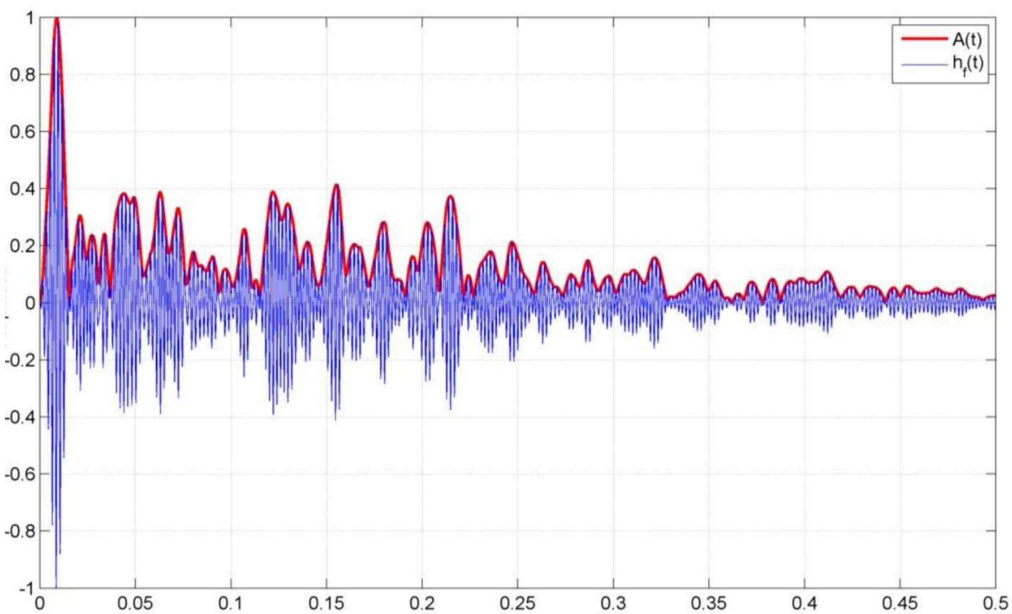
- To gain understanding of relative importance of low frequency temporal structure of speech and the effect of number of bands in speech recognition.

Theory:

In this experiment, the significance of the temporal and amplitude information of the audio signal in speech recognition is observed. To achieve this, firstly the envelope is extracted and then the signal is noise modulated to remove the spectral information.

<u>Extraction of envelope</u>-

To extract the envelope, Hilbert Transform is used which gives a complex valued signal. When its absolute value is taken, it follows the peaks of the original audio signal.



<u>Noise Modulation</u>-

The obtained envelope is multiplied to band-limited white noise to obtain the DSB-SC noise modulated signal. This signal is given by-

$$x(t) = e(t).wn(t)$$

Where x(t) is the noise-modulated signal, e(t) is the extracted envelope and wn(t) is band-limited noise.

Observations and results:

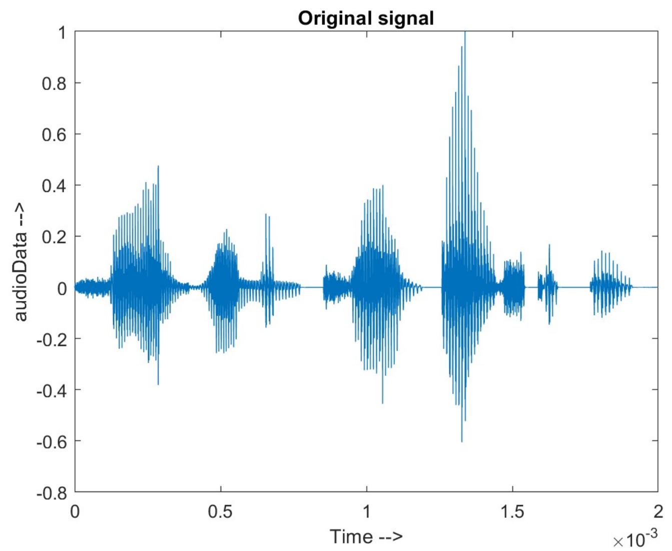
Code-

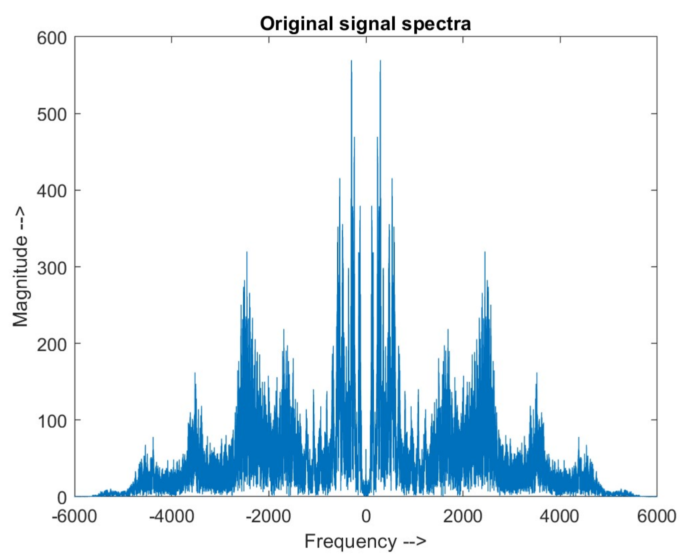
```
[audioData,fs] = audioread("fivewo.wav");
%soundsc(audioData,fs)
pow = bandpower(audioData);
pow = 20*log10(pow);
T = 0.002;
t = 0: (1/(800*fs)): T;
n = length(t);
fshift = (-n/2:n/2-1)*(fs/n);


plot(t, audioData)
title('Original signal')
xlabel('Time -->')
ylabel('audioData -->')


audio_fft = fftshift(fft(audioData));
audio_fft = abs(audio_fft);
plot(fshift,audio_fft);
xlim([-6000,6000]);
title('Original signal spectra')
xlabel('Frequency -->')
ylabel('Magnitude -->')


white_noise = wgn(length(audioData),1,pow);

%4 bands
z =  90*((2^1.5).^(0:4));%2 logarithmically spaced bandss
[b1,a1] = butter(2,[z(1)/(fs/2) z(2)/(fs/2)],'bandpass');
y1 = filter(b1,a1,audioData);
wn1 = filter(b1,a1,white_noise);
[b2,a2] = butter(2,[z(2)/(fs/2) z(3)/(fs/2)],'bandpass');
y2 = filter(b2,a2,audioData);
wn2 = filter(b2,a2,white_noise);
[b3,a3] = butter(2,[z(3)/(fs/2) z(4)/(fs/2)],'bandpass');
y3 = filter(b3,a3,audioData);
wn3 = filter(b3,a3,white_noise);
[b4,a4] = butter(2,[z(4)/(fs/2) z(5)/(fs/2)],'bandpass');
y4 = filter(b4,a4,audioData);
wn4 = filter(b4,a4,white_noise);


y_out1 = hilbert(y1);
yo1 = abs(y_out1);
y_out2 = hilbert(y2);
yo2 = abs(y_out2);
```

```
y_out3 = hilbert(y3);
yo3 = abs(y_out3);
y_out4 = hilbert(y4);
yo4 = abs(y_out4);


[bo,ao] = butter(4 ,240/(fs/2) ,'low'); %Low-pass filter
envelope1 = filter(bo, ao, yo1);%Detected message signal
envelope2 = filter(bo, ao, yo2);
envelope3 = filter(bo, ao, yo3);
envelope4 = filter(bo, ao, yo4);
final_envelope = envelope1 + envelope2 +envelope3 + envelope4;
envelopeb4 = final_envelope;


recons_audiop1 = envelope1.*wn1;
recons_audiop2 = envelope2.*wn2;
recons_audiop3 = envelope3.*wn3;
recons_audiop4 = envelope4.*wn4;


%Plotting envelopes
figure(5);
subplot(4,1,1)
plot(t, envelope1)
title('Envelope1')
xlabel('time -->')
ylabel('y -->')
subplot(4,1,2)
plot(t, envelope2)
title('Envelope2')
xlabel('time -->')
ylabel('y -->')
subplot(4,1,3)
plot(t, envelope3)
title('Envelope3')
xlabel('time -->')
ylabel('y -->')
subplot(4,1,4)
plot(t, envelope4)
title('Envelope4')
xlabel('time -->')
ylabel('y -->')


recons_audio = recons_audiop1+recons_audiop2+recons_audiop3+recons_audiop4;
figure(6);
plot(t,recons_audio)
title('Noise Modulated Audio for 4 bands')
xlabel('Time -->')
ylabel('Magnitude -->')
audio_fft = fftshift(fft(recons_audio));
```

```
audio_fft = abs(audio_fft);
plot(fshift,audio_fft);
xlim([-50000,50000]);
title('Final signal spectra for 4 bands')
xlabel('Frequency -->')
ylabel('Magnitude -->')


audiowrite('test4.wav',envelope,fs); %Saving results

%output
% [audio_out,fs] = audioread("test1.wav");
% soundsc(audio_out,fs)
```

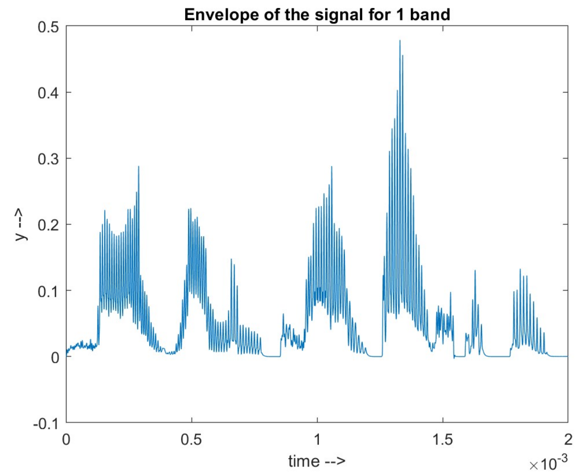
Plots-

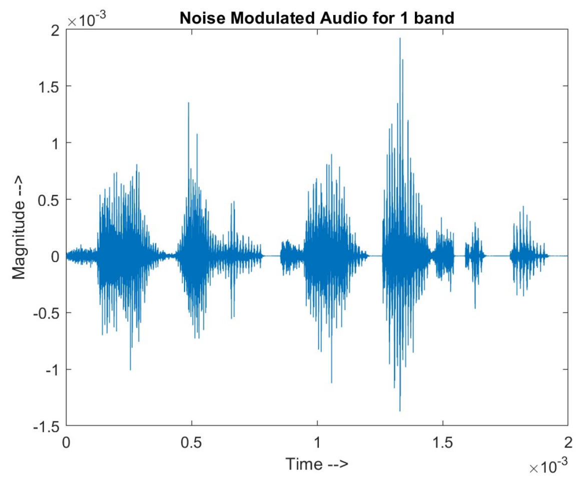


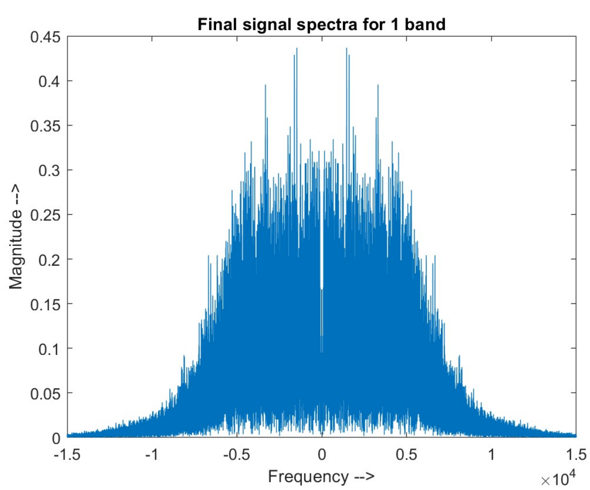


1 Band-

90Hz-5.76kHz

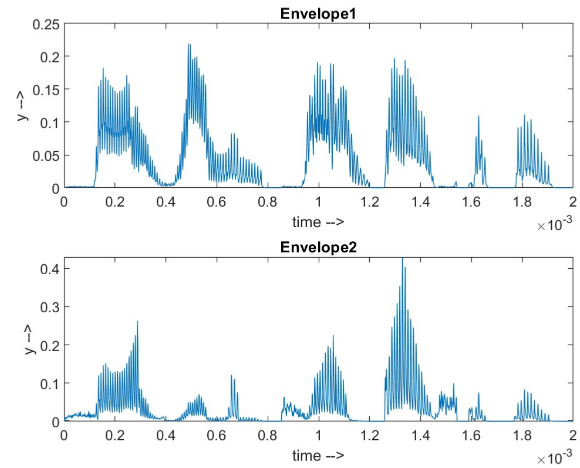


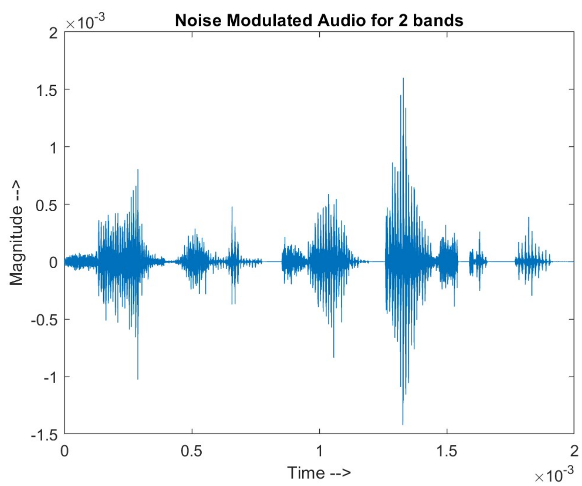


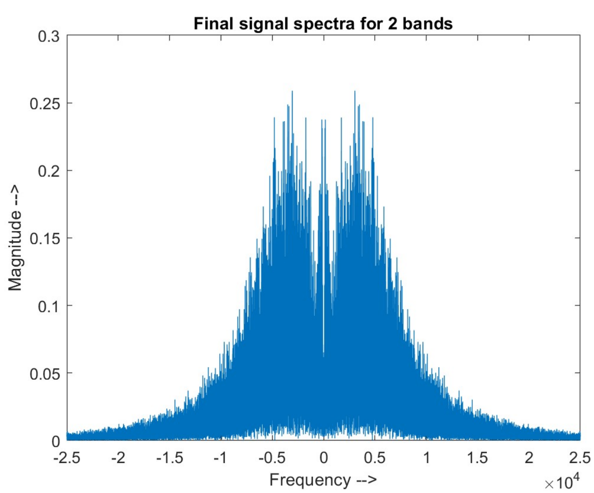


<u>2 Bands</u>-

90-720Hz, 720-5760Hz

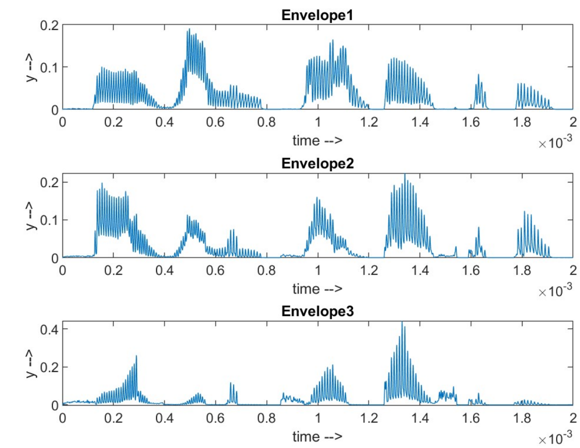


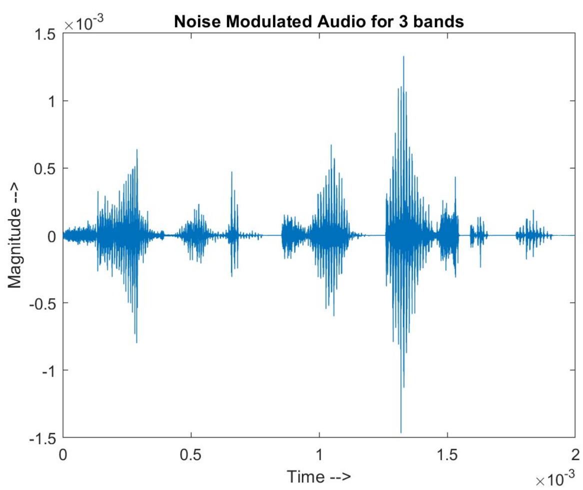


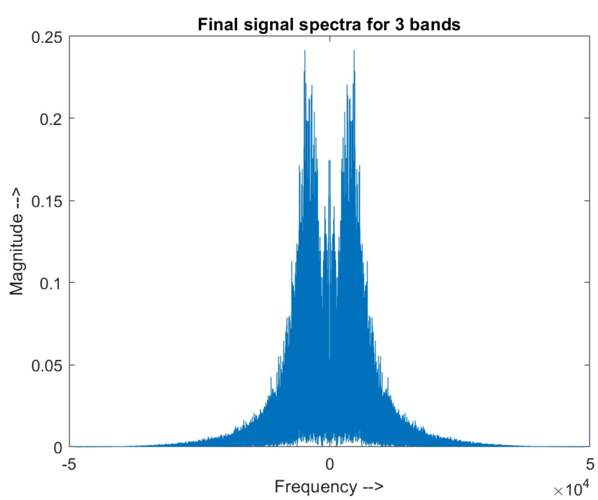


<u>3 Bands</u>-

90-360 Hz, 360-1440Hz, 1440-5760Hz

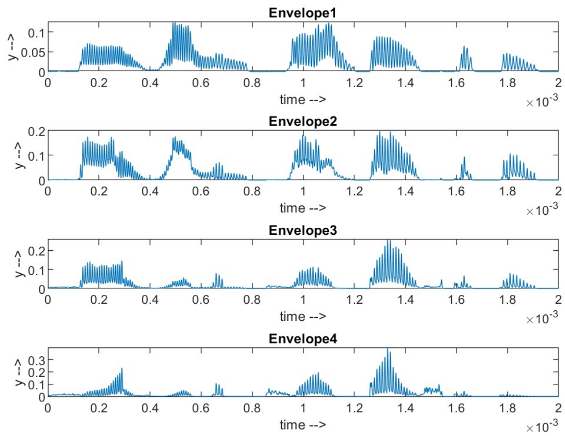


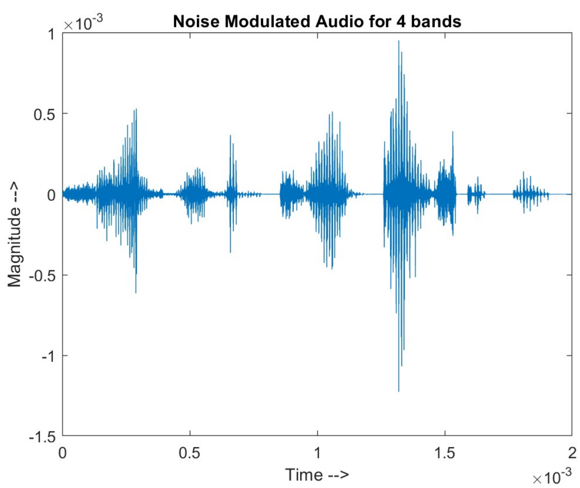


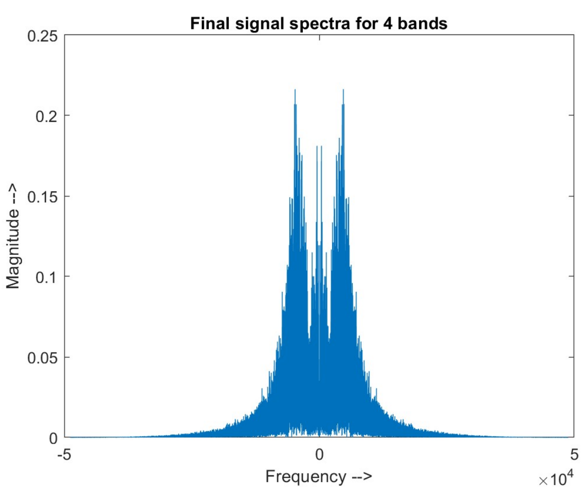


<u>4 Bands</u>-

90-254.56Hz, 254.56-720Hz, 720-2036.5Hz, 2036.5-5760Hz

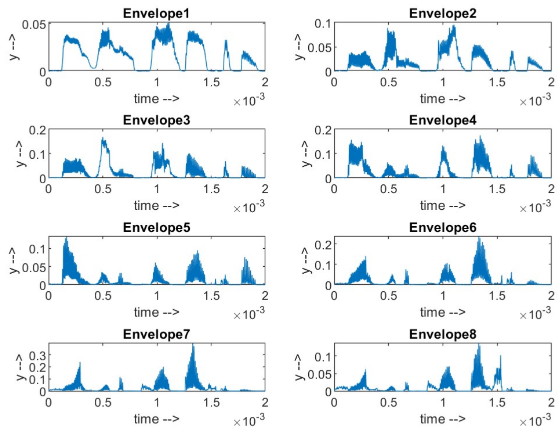


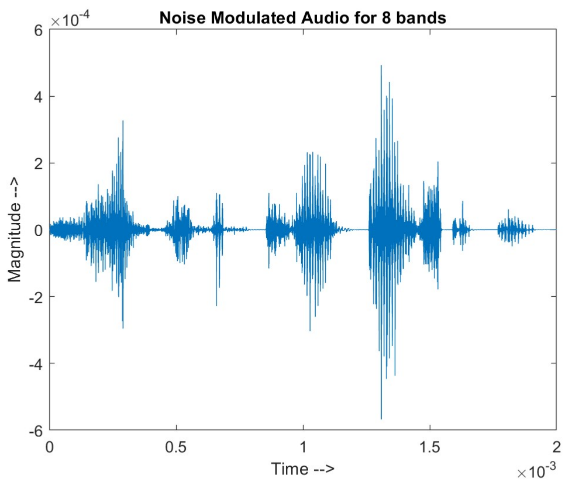


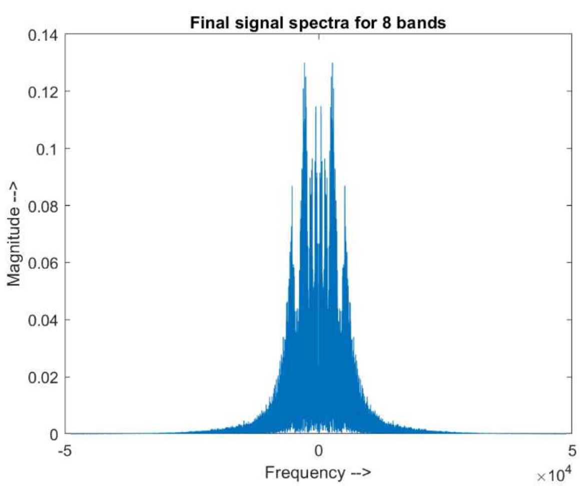


<u>8 Bands</u>-

90-151.3614Hz, 151.3614-254.56Hz, 254.56-428.11Hz, 428.11-720Hz, 720-1210.9Hz, 1210.9-2036.5Hz, 3424.9-5760 Hz

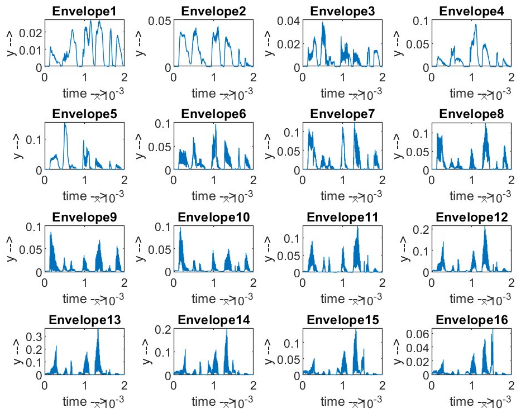


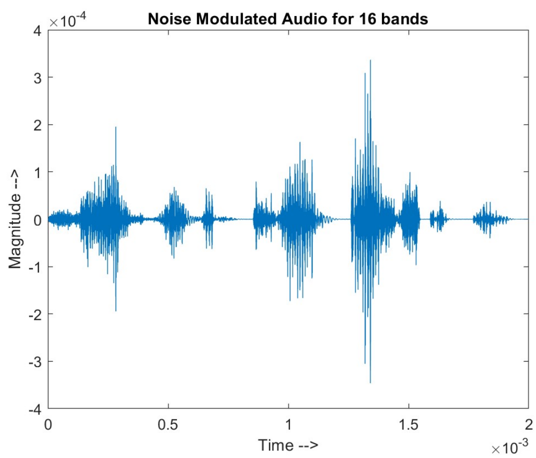


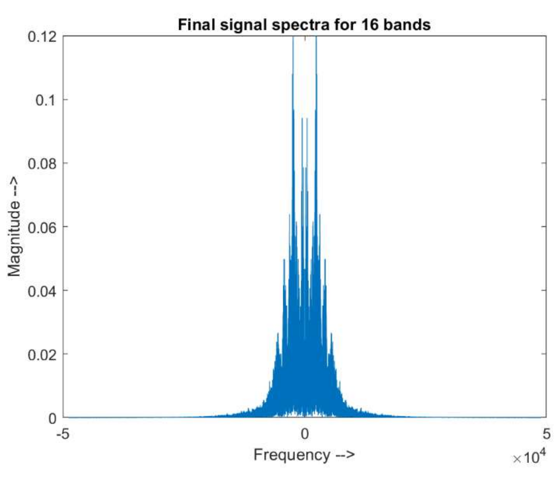


16 Bands-

90-116.71Hz, 116.71-151.36Hz, 151.36-196.29Hz, 196.29-254.56Hz, 254.56-330.12Hz, 330.12-428.11Hz, 428.11-555.19Hz, 555.19-720Hz, 720-933.72Hz, 933.72-1210.9Hz, 1210.9-1570.3Hz, 1570.3- 2036.5Hz, 2036.5-2641Hz, 2641-3424.9Hz, 3424.9-4441.6Hz, 4441.6-5760 Hz







Discussion:

- After noise modulation, it can be observed that the frequency spectra of the resultant signal does not resemble the frequency spectra of the original audio signal, especially for lesser number of bands. This is because most of the spectral information is removed when the audio signal is multiplied with white noise.

- The information stored in the amplitude is extracted by getting the envelope of the signal. This is done by getting the Hilbert Transform of the signals in different bands.

- As the number of bands is increased, the noise modulated signal looks closer the audio signal. The vocal patterns of the speech became closer to the original audio signal as the number of bands are increased.

- So it can be observed, that the articulation of the speech is mostly due to the spectral information of the signal as it is hard to hear the exact words spoken for the noise modulated signal but the rhythmic pattern of the speech can be recognized. This is because the amplitude and temporal information contain the vocal cues of the speech which can be crucial for the applications in speech recognition.